

Aplicación de la implicación textual en un sistema de la validación de respuestas automáticas

Partha Pakray¹, Alexander Gelbukh² y Sivaji Bandyopadhyay¹

¹ Jadavpur University, India

² Instituto Politécnico Nacional, México

www.parthapakray.com, www.gelbukh.com,
www.sivajibandyopadhyay.com

Resumen. La tarea de la respuesta automática a preguntas consiste en obtener de forma automática las respuestas a las preguntas del usuario basándose en una gran colección de documentos o la Internet abierta. Se diferencia de la tarea de recuperación de información en que su resultado es una o varias respuestas cortas y no un conjunto de documentos. Uno de los pasos en el proceso de respuesta automática a preguntas es la validación de la respuesta: la evaluación de la idoneidad de una posible respuesta para la pregunta específica. Dados dos textos llamados T (texto) y H (hipótesis), la tarea de reconocimiento de implicación textual consiste en decidir si el significado de H se puede deducir lógicamente del significado de T. Hemos aplicado dos técnicas diferentes de implicación textual a la tarea de la validación de la respuesta en el contexto de la tarea de respuesta automática a preguntas. En nuestros experimentos con el conjunto de prueba en inglés AVE 2008, se obtuvo la puntuación $c@1$ de 0.65 y 0.69, respectivamente.

Palabras clave: respuesta automática a preguntas, implicación textual, validación de la respuesta, reconocimiento de entidades nombradas.

An Application of Textual Entailment Recognition in an Answer Validation System

Abstract. The question answering task consists in automatically obtaining answers to the user's questions basing on a large collection of documents or the open Internet. It differs from the information retrieval task in that the result is one or several short answers and not a set of documents. One of the steps in the question answering process is answer validation: evaluation of the appropriateness of a candidate answer for the question. Given two texts called T (Text) and H (Hypothesis), the textual entailment recognition task consists in deciding whether the meaning of H can be logically inferred from that of T. We applied the two different textual entailment techniques to the answer validation task in the context of question answering. In our experiments on the AVE 2008 English test set, we obtained the $c@1$ score of 0.65 and 0.69, respectively.

Keywords: question answering, textual entailment, answer validation, named entity recognition.

1. Introducción

El reconocimiento de la implicación textual (RIT) [1] es uno de los desafíos recientes del Procesamiento del Lenguaje Natural (PLN) que tiene un amplio potencial de las aplicaciones en otras tareas del PLN tales como la generación automática de resúmenes, paráfrasis, extracción de información, traducción automática y respuesta automática a preguntas.

La tarea del RIT se define como una relación direccional entre pares de expresiones textuales, denotados T –el “texto”, lo implicante–, y H –la “hipótesis”, lo implicado. Si el significado de H se puede inferir a partir del significado de T entonces T implica H. La tarea de RIT consiste entonces en decidir automáticamente si tal relación existe o no entre los T y H dados.

Se han organizado seis competiciones internacionales sobre esta tarea: RTE-1 a RTE-7 (por las siglas en inglés, “*Recognizing Textual Entailment*”) en los años 2005 a 2011, respectivamente. El trabajo aquí presentado, en sus versiones anteriores, ha participado en las RTE-5 y RTE-6.

La tarea de la validación de la respuesta está dirigida a la creación de los sistemas que miden la exactitud de la respuesta a una pregunta, dado un texto de soporte. Específicamente, tales sistemas reciben un conjunto de tripletas: pregunta, respuesta y el texto de soporte, y para cada tal tripleta devuelven una de las tres posibles respuestas: “seleccionado”, “validado” o “rechazado”. Aquí las respuestas rechazadas

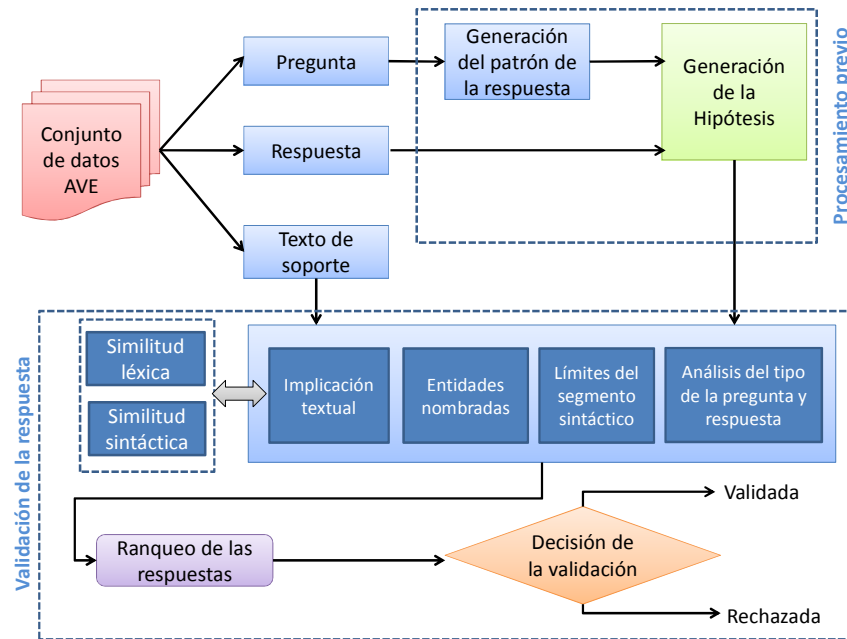


Figura 1. Arquitectura del sistema de la validación de respuestas

```

<q id="0061" leng="EN">
  <q_str>¿Dónde nació Joseph Fourier?</q_str>
  <a id="0061_1" valor="rechazado">
    <a_str>Paris</a_str>
    <t_str doc="Joseph Fourier">Joseph Fourier Joseph Fourier Joseph Fourier Jean
    Baptiste Joseph Fourier Nacido 21 de marzo 1768 Auxerre, Yonne, France Fallecido el 16
    de mayo 1830 París, Francia Residencia Francia Nacionalidad francesa Campo
    Matemático, físico e historiador Institución École Normale École Polytechnique Alma
    Mater École Normale Asesor Doctoral Joseph Lagrange Estudiantes de doctorado Gustav
    Dirichlet Giovanni Plana Conocido por transformada de Fourier Religión Católica Jean
    Baptiste Joseph Fourier (21 marzo 1768 — 16 mayo 1830) fue un matemático y físico
    francés quien es mejor conocido por iniciar la investigación de la serie de Fourier y sus
    aplicaciones a los problemas de flujo de calor.</t_str>
  </a>
  <a id="0061_2" valor="desconocido">
    <a_str>Francia</a_str>
    <t_str doc="Joseph Fourier">Vida Fourier nació en Auxerre, en el
    departamento de Yonne de Francia, hijo de un sastre.</t_str>
  </a>
  <a id="0061_3" valor="rechazado">
    <a_str>Grenoble</a_str>
    <t_str doc=" Universidad Joseph Fourier">Universidad Joseph Fourier Universidad
    Joseph Fourier Université Joseph Fourier (Universidad Joseph Fourier) es una universidad
    francesa, situada en la ciudad de Grenoble, centrada en los campos de las ciencias, las
    tecnologías y la salud.</t_str>
  </a>
  <a id="0061_8" valor="rechazado">
    <a_str>Auxerre</a_str>
    <t_str doc="Auxerre.html">Auxerre.</t_str>
  </a>
  ....
</q>

```

Figura 1. Traducción de un ejemplo de un documento del conjunto de documentos del ejercicio de validación de respuesta (AVE 2008).

son las que no son apropiadas para la pregunta, validadas son apropiadas, y una de las validadas se elige como seleccionada, y será ésta la que será devuelta al usuario. Se han organizado tres competiciones internacionales en esta tarea: AVE 2006 [2] (por sus siglas en inglés, “*Answer Validation Exercise*”), AVE 2007 [3] y AVE 2008 [4].

Hemos desarrollado un sistema de implicación textual en el escenario monolingüe y el de idiomas diferentes. Estamos aplicando nuestra técnica del RIT en los sistemas de respuesta a preguntas y de validación de la respuesta. Para la tarea de RIT hemos desarrollado tanto sistemas basados en reglas como sistemas basados en el aprendizaje automático que tienen en cuenta los rasgos semánticos, sintácticos y léxicos. Nos dirigimos principalmente a establecer la relación de implicación textual (IT) basándonos en las medidas de similitud léxica, sintáctica y semántica.

Tabla 1. Detección del límite del segmento sintáctico (traducción literal del inglés).

| | | |
|------------------|--|--|
| Ejemplo 1 | Pregunta | ¿Qué fue la nacionalidad de Jacques Offenbach? |
| | Analizador de categorías gramaticales de Stanford | ¿/. Qué/WP fue/VBD la/DT nacionalidad/NN de/IN Jacques/NNP Offenbach/NNP ?/. |
| | Analizador de segmentos sintácticos con los campos aleatorios condicionales | Qué/WP/B-NP fue/VBD/B-VP la/DT/B-NP nacionalidad/NN/I-NP de/IN/B-PP Jacques/NNP/B-NP Offenbach/NNP/I-NP |
| | Límite del segmento | (Qué/WP/B-NP) (fue/VBD/B-VP) (la/DT/B-NP#nacionalidad/NN/I-NP) (de/IN/B-PP) (Jacques/NNP/B-NP#Offenbach/NNP/I-NP#) |
| Ejemplo 2 | Pregunta | ¿Dónde fue Joseph Fourier nacido? |
| | Analizador de categorías gramaticales de Stanford | ¿/. Dónde/WRB fue/VBD Joseph/NNP Fourier/NNP nacido/VBN ?/. |
| | Analizador de segmentos sintácticos con los campos aleatorios condicionales | Dónde/WRB/B-ADVP fue/VBD/B-VP Joseph/NNP/B-NP Fourier/NNP/I-NP nacido/VBN/B-VP |
| | Límite del segmento | (Dónde/WRB/B-ADVP) (fue/VBD/B-VP) (Joseph/NNP/B-NP#Fourier/NNP/I-NP) (nacido/VBN/B-VP#) |

Para eso, hemos usado la detección de las entidades nombradas, categorías gramaticales, similitud basada en el diccionario WordNet, similitud basada en segmentos sintácticos (*chunks* en inglés), similitud basada en n-gramas, similitud basada en la distancia léxica, similitud sintáctica, similitud semántica y otras para decidir si existe la relación de IT con aproximaciones basadas en reglas y en el aprendizaje de máquina, tratando el problema de RIT como un problema de clasificación.

Con estos desarrollos, hemos participado en diferentes competencias internacionales de evaluación que se concentraron principalmente en las cuestiones de la clasificación en dos vías y la clasificación en vías múltiples, en ambos casos tanto en el escenario monolingüe como en el escenario de idiomas distintos (*cross-lingual* en inglés).

En la tarea de RIT de la competencia principal de la Conferencia de Análisis de Texto (TAC, por sus siglas en inglés) en su quinta edición (2009) propusimos un sistema basado en léxico [5]. En la tarea RIT de la competencia principal de la TAC en su sexta edición (2010) propusimos un sistema basado en la similitud léxica y sintáctica [6]. Hemos desarrollado otro sistema [7] que se basa en la evaluación de similitud entre pares T-H en lugar de las etiquetas de implicación.

Hemos utilizado el Lenguaje de Redes Universal (UNL, por sus siglas en inglés: Universal Networking Language, www.undl.org) para identificar los rasgos semánticos. UNL tiene todos los componentes de un lenguaje natural. Hemos desarrollado un sistema de RIT basado en UNL que compara las relaciones de UNL en el texto y en la hipótesis.

Tabla 2. Regla de la generación del patrón para el tipo de pregunta “¿Qué” (traducción literal del inglés).

| Pregunta tipo “¿Qué” | | |
|----------------------|--|---|
| Regla 1 | ¿Qué VP NP? | ¿Qué es astronomía? |
| Patrón | NP VP </Respuesta> | Astronomía es </Respuesta> |
| Regla 2 | ¿Qué NP VP NP? | ¿Qué vehículo es usado en el deporte “Moto cross”? |
| Patrón | </Respuesta> VP NP | </Respuesta> es usado en el deporte “Moto cross” |
| Regla 3 | ¿Qué <ser-verbo> <el nombre de>/<la nacionalidad de>/<la ocupación de> NP? | ¿Qué fue la nacionalidad de Jacques Offenbach? |
| Patrón | <el nombre de> NP <ser-verbo> </Respuesta> | La nacionalidad de Jacques Offenbach fue </Respuesta> |
| Regla 4 | ¿<PP> qué NP <ser-verbo> NP VP? | ¿En qué países es la navegación deportiva sobre hielo practicado? |
| Patrón | NP <ser-verbo> VP PP </Respuesta> | La navegación deportiva sobre hielo es practicada en </Respuesta> |
| Regla 5 | ¿PP qué NP1 <ser-verbo> NP2 VP? | ¿En qué año fue Emerson Lake & Palmer formada? |
| Patrón | NP2 VP PP </Respuesta> | Emerson Lake & Palmer se formó en </Respuesta> |
| Regla 6 | ¿PP qué NP1 <ser-verbo> que NP2 VP NP3? | ¿En qué año fue que James Hurt ganó el mundial de Formula Uno? |
| Patrón | NP2 VP NP3 PP </Respuesta> | James Hurt ganó el mundial de Formula Uno en </Respuesta> |
| Regla 7 | ¿PP que NP1 <ser-verbo> NP2 VP? | ¿En qué año será el Hotel Brunswick City en Glasgow abierto? |
| Patrón | NP2 <ser-verbo> VP PP </Respuesta> | El Hotel Brunswick City en Glasgow será abierto en </Respuesta> |

En este artículo presentamos nuestro siguiente paso: un sistema de validación de respuesta [8] basado en la RIT. Primero, combinamos la pregunta y la respuesta en la hipótesis H, y consideramos el texto de soporte como el texto T, para comprobar la relación de IT como validada o rechazada. Hemos aplicado la técnica de IT al sistema de validación de las respuestas para tomar decisiones de validación o rechazo. También hemos aplicado la técnica de IT en las tareas de respuesta automática a preguntas (por ejemplo, QA4MRE) en CLEF 2011 [9] y CLEF 2012 [10].

2. La arquitectura de nuestro sistema de validación de respuestas

Hemos construido un sistema de validación de respuesta mostrado en la figura 1, donde la validación se basa en la IT. Las características importantes que se utilizaron para el desarrollo del sistema son la generación del patrón, la generación de la

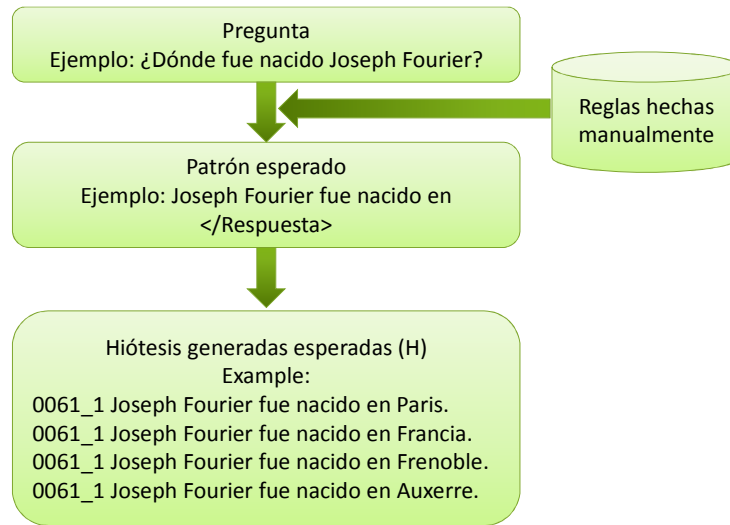


Figura 2. Un ejemplo de la generación de la hipótesis.

hipótesis, RIT, el reconocimiento de entidades nombradas, análisis de tipo de la pregunta-respuesta, el módulo de límites del segmento sintáctico (*chunk*) y el módulo de la similitud sintáctica. El sistema está basado en reglas. En primer lugar, combinamos la pregunta y la respuesta en la Hipótesis (H), y el texto de soporte lo consideramos como el Texto (T), con lo cual obtenemos las decisiones de la IT como validación o rechazado.

2.1. Conjunto de datos

El experimento se ha llevado a cabo principalmente con el conjunto de datos AVE 2008. Un ejemplo de este conjunto de datos se muestra en la figura 2.

El sistema analiza la pregunta usando el etiquetador de las categorías gramaticales de Stanford (<http://nlp.stanford.edu/software/tagger.shtml>) y el analizador de segmentos sintácticos (*chunker*) basado en los campos aleatorios condicionales (CRF, por sus siglas en inglés, <http://crfchunker.sourceforge.net>). Después de este procesamiento, el sistema detecta los límites de los segmentos sintácticos, como se muestra en la tabla 1.

En este módulo, cada pregunta es convertida en un patrón de respuesta afirmativa en la ranura de la respuesta (</Respuesta>). Las reglas para el patrón de respuesta correspondiente a cada pregunta fueron desarrollados hecha manualmente. La generación de un patrón para la pregunta de tipo “¿qué?” (“what” en inglés) se muestra en la tabla 2.

Tabla 3. Respuesta sustituida por la regla de generación de patrones.

| N del par | Hipótesis generadas esperadas |
|-----------|--|
| 0061_1 | Joseph Fourier fue nacido en Paris. |
| 0061_2 | Joseph Fourier fue nacido en Francia. |
| 0061_3 | Joseph Fourier fue nacido en Grenoble. |
| 0061_8 | Joseph Fourier fue nacido en Auxerre. |

2.3. Módulo de la generación de la hipótesis

Después de la generación del patrón, la plantilla `</Respuesta>` se sustituye por la cadena de la respuesta (`<a_str>`) que forma la hipótesis generada. Ahora, el sistema tiene el Texto (T), el texto de apoyo y la Hipótesis (H), la hipótesis generada. Por ejemplo, para la pregunta número 0061 en el conjunto de prueba AVE-2008, el sistema genera las hipótesis para cada una de las respuestas alternativas como se describe en la tabla 3.

Un ejemplo de la generación de la hipótesis se muestra en la figura 3.

2.4. Módulo del reconocimiento de la implicación textual

El módulo más importante de nuestro sistema es el módulo de la IT. El proceso de RIT se basa en la implicación léxica y sintáctica, es decir, es un proceso de RIT denominado híbrido. En esta sección describimos un sistema de RIT de dos vías (validación o rechazo) que utiliza las características léxicas y sintácticas. La arquitectura del sistema se muestra en la figura 4. Nuestro sistema híbrido de RIT utiliza el aprendizaje basado en la máquina de soporte vectorial, con treinta y cuatro rasgos usados para el entrenamiento: cinco rasgos del proceso del RIT léxico, diecisiete rasgos de la distancia léxica y once rasgos del sistema del RIT de dos vías basado en la información sintáctica. Los cinco rasgos léxicos se describen en detalle en [4] y los once rasgos sintácticos se describen en detalle en [11].

Los rasgos de similitud léxica que se utilizan en el sistema actual para comparar la hipótesis con el texto son:

- comparación de unigramas basada en WordNet,
- comparación de bigramas,
- subsecuencia común más larga,
- salta-gramas (skip-grams),
- comparación de las bases morfológicas.

Los rasgos de la distancia léxica, véanse los detalles en [12], son los siguientes:

- distancia de bloque,
- derivación de la longitud de Chapman,
- longitud media de Chapman,
- similitud del coseno,
- similitud de Dice,
- distancia euclidiana,
- similitud de Jaccard,
- distancia de Jaro,
- distancia de Jaro-Winkler,
- coeficiente de coincidencia,
- distancia de Monge-Elkan,
- distancia de Needleman-Wunch,

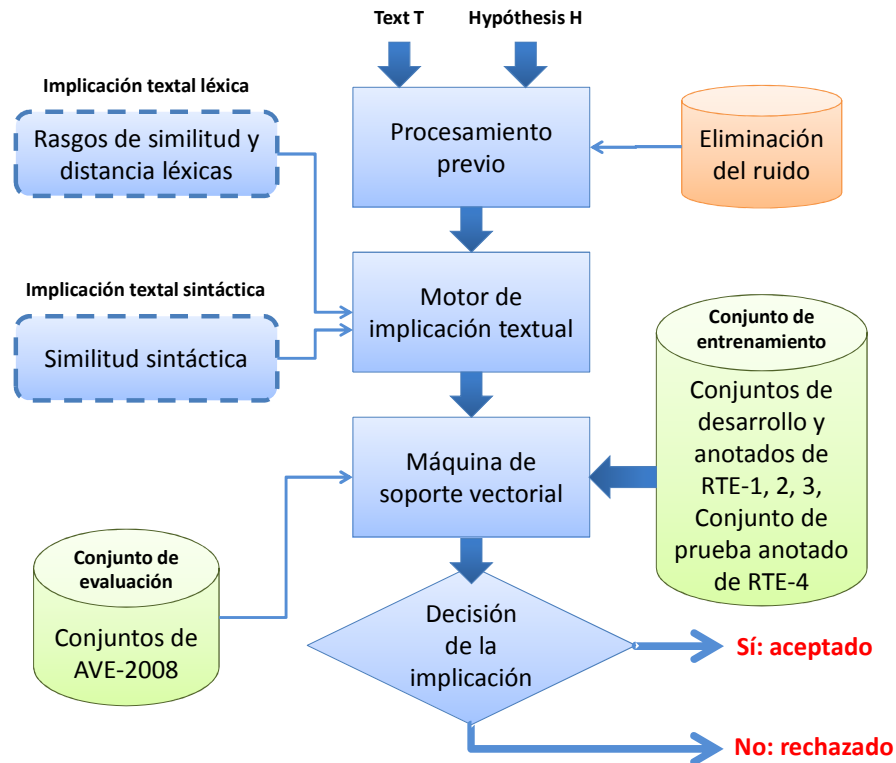


Figura 3. La arquitectura del sistema de reconocimiento de la implicación textual.

- coeficiente de traslape,
- distancia en qgramas,
- distancia de Smith-Waterman,
- distancia de Smith-Waterman con ir-a,
- distancia de Smith-Waterman con ir-a de ventanas afina.

Para el RTE basado en la información sintáctica, se usaron los siguientes rasgos:

- comparación del sujeto con el sujeto,
- comparación del sujeto y el verbo basada en WordNet,
- comparación del sujeto y el objeto,
- comparación del objeto y el verbo,
- comparación del objeto y el verbo basada en WordNet,
- comparación del sujeto y el objeto cruzada,
- comparación del número,
- comparación de los sustantivos,
- comparación de las frases preposicionales,
- comparación de los determinantes,
- comparación de otras relaciones.

Tabla 4 Entidades nombradas reconocidas por la herramienta LT-TTT2 (traducción literal del inglés).

| Para la entidad nombrada | Etiqueta dada por la herramienta LT-TTT2 | Ejemplo |
|--------------------------|--|--------------------------|
| UBICACIÓN | <enamex tipo="ubicación"> | India, Francia etc. |
| ORGANIZACIÓN | <enamex tipo="organización"> | BBVA Bancomer etc. |
| PERSONA | <enamex tipo="persona"> | Henry Purcell etc. |
| FECHA | <timex tipo="fecha"> | 24 de junio de 1982 etc |
| TIEMPO | <timex tipo="tiempo"> | La tarde etc |
| PORCENTAJE | <numex tipo="porcentaje"> | 5% etc |
| DINERO | <numex tipo="dinero"> | \$7 millones etc |
| PAPEL | <papel> | Capitán, Presidente etc. |
| NÚMERO | <phr c="cd"> | Dos, 6 etc |

El sistema utiliza la biblioteca LIBSVM (<http://www.csie.ntu.edu.tw/~cjlin/libsvm>) para construir el archivo del modelo. El sistema del RIT utiliza los siguientes conjuntos de datos: el conjunto de desarrollo y de prueba de la RTE-1, los conjuntos de desarrollo y los conjunto de prueba anotados de las RTE-2, RTE-3 y el conjunto de prueba anotado de la RTE-4 para la tarea de la clasificación de dos vías para el entrenamiento para construir el archivo del modelo. La LIBSVM utiliza un clasificador para aprender de este conjunto de datos. Para el entrenamiento fueron utilizados 3967 pares de texto y hipótesis. Después del entrenamiento del sistema lo evaluamos en el conjunto de datos de AVE. Como el resultado, el sistema produce las decisiones del RIT: “sí” –validar– o “no” –rechazar–.

2.5. Módulo del reconocimiento de las entidades nombradas

Este módulo se basa en la detección y comparación de las entidades nombradas en el par del texto de soporte (T) y la hipótesis (H) generada. Una vez se han detectados las entidades nombradas de la hipótesis y el texto, el siguiente paso es determinar el número de las entidades nombradas en la hipótesis que coinciden con las del texto correspondiente. La medida de coincidencia NE_Match se define como sigue:

$$NE_Match = \frac{\text{número de entidades nombradas en común en T y H}}{\text{número de entidades nombrados en la H}}$$

Si el valor de NE_Match es 1, entonces todas las entidades nombradas en la hipótesis coinciden con las en el texto. En este caso el par de texto y hipótesis se considera como un caso de implicación y al par T y H se le asigna el valor “validado”. De lo contrario, a este par se le asigna el valor “rechazado” con una puntuación de coincidencia.

Para el reconocimiento de las entidades nombradas, usamos el llamado componente de entidad nombrada LT-TTT2 (<http://www.ltg.ed.ac.uk/software/lt-ttt2>). Este componente de software reconoce y marca las siguientes clases de las entidades nombradas: expresiones numéricas (por ejemplo, cantidades de dinero y porcentajes), expresiones temporales (por ejemplo, fechas y horas), expresiones de entidad

Tabla 5. Tipo de respuesta esperada por el tipo de pregunta.

| Tipo de pregunta | Respuesta esperada |
|------------------|--------------------|
| ¿Quién | PERSONA |
| ¿Cuándo | FECHA / TIEMPO |
| ¿Dónde | UBICACIÓN |
| ¿Qué | OBJETO |
| ¿Cómo | MEDIDA |

nombrada (por ejemplo, personas, organizaciones y lugares) y papeles (por ejemplo, *presidente*).

Un ejemplo de la entidad designada por la herramienta el LT-TTT2 se muestra en la tabla 4.

2.6. Módulo del análisis del tipo de pregunta y respuesta

Las preguntas iniciales son preprocesadas utilizando el analizador sintáctico de dependencias de Stanford (<http://nlp.stanford.edu/software/lex-parser.shtml>). El tipo de la pregunta y el tipo de la respuesta esperada se identifican generalmente por la palabra principal de la pregunta, por ejemplo “¿qué”, “¿quien”, “¿dónde”, “¿cuándo”, etc. La tabla 5 muestra preguntas y el tipo de la respuesta correspondiente que se espera.

Por ejemplo, si el tipo de pregunta es “¿donde”, el tipo de la respuesta que se espera es una *ubicación*. La cadena de respuesta “<a_str>” es analizada sintácticamente por el componente LT-TTT2 antes mencionado. Si el componente LT-TTT2 genera la etiqueta “<entidad_nombrada tipo=ubicación>”, entonces la cadena de respuesta que es validada, de lo contrario es rechazada. Un ejemplo se muestra en la tabla 6.

Para el tipo “¿qué” buscamos la palabra clave (por ejemplo, una *empresa*) relacionada con la palabra “¿qué” a través de una relación de dependencia. Si la palabra clave es una *empresa*, entonces el tipo de la respuesta esperada es “organización”. Si la cadena de la respuesta correspondiente está marcada por la herramienta LT-TTT2 como “<entidad_nombrada tipo=organización>”, entonces la cadena de la respuesta se marca como validada, de lo contrario, como rechazada.

Si el tipo de pregunta es “¿cómo” y la cadena de respuesta se marcada como “<frase c=cd>” por la herramienta LT-TTT2, entonces la cadena de respuesta se marca como validada, de lo contrario, como rechazada.

Para cada comparación, se genera una puntuación cuantitativa. Un ejemplo de la validación a través del reconocimiento de las entidades nombradas se muestra en la tabla 7. Por último, el sistema otorga una puntuación final de coincidencia.

2.7. Módulo de la detección de los segmentos sintácticos (chunks)

Las oraciones con las preguntas son preprocesados utilizando el analizador sintáctico de dependencias de Stanford. Las palabras junto con su la información de la

Tabla 6. Análisis con la herramienta LT-TTT2.

| | |
|----------------------------|--|
| Pregunta | ¿Dónde la Olimpiada de Invierno tendrá lugar en 1994? |
| Respuesta esperada | UBICACIÓN |
| Hipótesis | La Olimpiada de Invierno de 1994 tendrá lugar en Lillehammer. |
| Hipótesis analizada | <code><timex tipo="fecha">Invierno</timex></code> <code><timex tipo="fecha">1994</timex></code> <code><enamet tipo="ubicación">Lillehammer</enamet></code> |

categoría gramatical se pasan al analizador de los segmentos sintácticos (*chunker*) basado en un campo aleatorio condicional (*English CRF Chunker*) para extraer de las preguntas los segmentos sintácticos de nivel de frase, por ejemplo, frase nominal (NP), frase verbal (VP), etc. Hemos desarrollado un módulo basado en reglas para identificar los límites de las frases. Las frases clave son identificadas para cada pregunta. Las frases que se relacionan con cada preposición constituyen la frase clave que corresponde a la relación de preposición. Cada frase verbal presente en la pregunta es también una frase clave. Si no hay relaciones preposicionales en la pregunta, entonces todas las frases son consideradas como frase clave, excepto la frase que forma la pregunta: “¿qué”, “¿dónde”, etc. Estas frases clave se buscan en el texto de soporte asociado con la oración de la pregunta. Un ejemplo de una pregunta con la información sobre los límites del segmento sintáctico (frase), las relaciones de dependencia y un conjunto de frases clave extraídos se muestran en la tabla 8.

El texto de soporte se analiza con el analizador sintáctico de dependencias de Stanford. La salida del analizador se pasa al analizador de los segmentos sintácticos basado en un campo aleatorio condicional (CRF) para extraer los segmentos de nivel de frase para cada oración en el texto de soporte. Las frases clave identificadas en la pregunta están luego comparados con las del texto de soporte asociado a la pregunta. Si un segmento clave (frase) completo coincide, heurísticamente el peso asignado a la coincidencia se define como sigue:

$$W_f = \frac{\text{longitud del segmento} + K}{\text{longitud del texto}},$$

donde K es el número de los segmentos (frases), para dar mayor peso a la coincidencia completa de los segmentos. Si existe una coincidencia parcial de tales segmentos, el peso se define sigue:

$$W_p = \frac{\text{longitud parcial del segmento}}{\text{longitud del texto}}.$$

El peso de un par de pregunta y el texto de soporte se identifica como el total de los pesos de los segmentos (frases) clave. Los pares de pregunta y el texto de soporte que reciben el peso máximo se identifican y las respuestas correspondientes se etiquetan como validados. If el par de la pregunta y el texto de soporte recibe un peso cero, entonces es etiquetado como rechazado, con una puntuación de coincidencia.

Tabla 7. Validación del reconocimiento de las entidades nombradas.

| Pregunta | ¿Dónde fue construido el Volkswagen Polo Playa? | | |
|-----------|---|------------------|------------------------------|
| Respuesta | UBICACIÓN | | |
| N del par | Respuesta | Entidad nombrada | Calificación de la respuesta |
| 0108_1 | Sudáfrica | Ubicación | validada |
| 0108_2 | Sucesor | | rechazada |
| 0108_3 | Apuntador | | rechazada |
| 0108_9 | Tom Ennis | Persona | rechazada |
| 0108_11 | Ibiza | | rechazada |
| 0108_12 | Playa | | rechazada |
| 0108_13 | Eurona | Ubicación | validada |

2.8. Módulo de la calificación cuantitativa de las respuestas

En este módulo, el sistema ya tiene los pesos generados por el módulo de RIT, el módulo del reconocimiento de las entidades nombradas, el módulo de análisis del tipo de la pregunta y el módulo de la detección de los límites de los segmentos sintácticos.

2.8.1. Módulo de la decisión sobre la validación de la respuesta

En ese módulo, el sistema finalmente decide cuál de las respuestas es correcta. Para la identificación de la respuesta correcta el sistema utiliza las siguientes dos técnicas.

Técnica 1 En esta variante, el sistema se usa una técnica de votación. Primero, el sistema compruebe si alguna entidad nombrada está presente en la hipótesis H generada. Se verifican las siguientes condiciones:

1. Si en la hipótesis H generada está presente una entidad nombrada, entonces el sistema comprueba el resultado del módulo del RIT de las entidades nombradas. Si este módulo indica “validado”, entonces se comprueban los resultados del módulo de la implicación textual, el módulo del análisis del tipo de la pregunta y respuesta, y el módulo de la segmentación sintáctica. Si todos estos módulos indican “validado”, entonces la respuesta dada a la pregunta inicial se marca como validada. De lo contrario, la respuesta se etiqueta como rechazada.
2. Si en la hipótesis H generada, no existen entidades nombradas, entonces se comprueban los resultados del módulo del RIT, módulo de análisis del tipo de la pregunta y respuesta, módulo de análisis de los segmentos sintácticos y el módulo de la similitud sintáctica. Si todos estos módulos generan el resultado “validado”, entonces esta respuesta a la pregunta inicial se marca como validada, de lo contrario, la respuesta es etiquetada como rechazada.

Tabla 8. Extracción de los segmentos sintácticos (*chunks*, traducción literal del inglés).

| Pregunta | ¿Qué fue la nacionalidad de Jacques Offenbach? |
|-----------------------------|--|
| Límite del segmento (frase) | (Qué/WP/B-NP) (fue/VBD/B-VP) (la/DT/B-NP nacionalidad/NN/I-NP) (de/IN/B-PP) (Jacques/NNP/B-NP Offenbach/NNP/I-NP) |
| Dependencias | [attr fue Qué det nacionalidad la nsubj fue nacionalidad nn Offenbach Jacques prep_of nacionalidad Offenbach] |
| Frase clave extraída | (fue/VBD/B-VP) (la/DT/B-NP nacionalidad/NN/I-NP) (Jacques/NNP/B-NP Offenbach/NNP/I-NP) |

Técnica 2 Para cada texto e hipótesis, el sistema calcula la puntuación de los módulos del RIT, de análisis de tipo de pregunta y respuesta, de la implicación a través de las entidades nombradas y de la similitud de los segmentos sintácticos. A continuación, el sistema considera la puntuación más alta para validar, y los pares restantes son etiquetados como rechazados.

3. Resultados experimentales

Hemos experimentado con nuestro sistema de validación de respuestas usando el conjunto de desarrollo para inglés del AVE 2008. Este conjunto consiste de 195 pares, de los cuales sólo 21 son positivos (10.77% del número total de los pares). Los valores del recuerdo (*recall*), precisión y la medida F obtenidos sobre este corpus para las respuestas marcadas como validadas se muestran en la tabla 9.

El conjunto anotado de prueba para el inglés del AVE 2008 consiste de 1055 pares, y el número de las respuestas marcadas como validadas es 79 (7.5% del total). Los valores del recuerdo (*recall*), precisión y la medida F obtenidos sobre este corpus para las respuestas marcadas como validadas se muestran en la tabla 10.

4. Conclusiones y trabajo futuro

La respuesta automática a preguntas es una nueva forma de la recuperación de información, mucho más cómoda para el usuario y mucho más apropiada para la comunicación humano-computadora [13]. La validación de la respuesta es un paso importante en el proceso de la respuesta automática a preguntas. Con este trabajo hemos demostrado que las técnicas del reconocimiento de la implicación textual (RIT)

Tabla 9. Conjunto de datos de desarrollo del AVE

| Conjunto de datos de desarrollo del AVE | Resultados con Técnica 1 | Resultados con Técnica 2 |
|---|--------------------------|--------------------------|
| “validado” en el conjunto de desarrollo | 21 | 21 |
| “validado” por nuestro sistema | 34 | 32 |
| coincidencia en la marca “validado” | 15 | 18 |
| Precisión | 0.44 | 0.56 |
| <i>Recall</i> | 0.71 | 0.85 |
| Medida F | 0.54 | 0.68 |

Tabla 10. Conjunto de datos de prueba del AVE.

| Conjunto de datos de prueba del AVE | Resultados con Técnica 1 | Resultados con Técnica 2 |
|-------------------------------------|--------------------------|--------------------------|
| “validado” en el conjunto de prueba | 79 | 79 |
| “validado” por nuestro sistema | 78 | 75 |
| coincidencia en la marca “validado” | 52 | 54 |
| Precisión | 0.66 | 0.72 |
| <i>Recall</i> | 0.65 | 0.68 |
| Medida F | 0.65 | 0.69 |

se pueden usar con éxito para la validación y ponderación de las posibles respuestas en el contexto de la respuesta automática a preguntas.

Hemos presentado en detalle la arquitectura del sistema de validación de la respuesta, la cual consiste en varios módulos desde el procesamiento previo de los datos hasta el reconocimiento de la implicación textual y la toma final de la decisión. Los resultados experimentales confirman nuestra hipótesis en cuanto a la utilidad de los métodos de RIT para la validación de la respuesta.

En nuestro trabajo futuro, planeamos experimentar con otros rasgos para el aprendizaje automático, así como con representaciones del texto más ricas, tales como los grafos conceptuales, n-gramas sintácticos [14] etc. Finalmente, intentaremos aplicar a nuestro trabajo las recientes técnicas de extracción abierta de los hechos [15] los cuales, como pensamos, pueden jugar el papel de los segmentos sintácticos pero mucho mejor. Esta generalización es aún más importante dado que existen sistemas para la extracción de los hechos no sólo para el inglés, sino también para el español [16].

Agradecimientos. Este trabajo fue parcialmente patrocinado por el Gobierno de México y el Gobierno de la India a través del proyecto DST India – CONACYT México 122030 / Mexico-India 2011–2014 “*Answer Validation through Textual Entailment*”, la Unión Europea a través del 7o Programa Marco, proyecto 269180 “*Web Information Quality Evaluation Initiative*” (WIQ-EI) y el Gobierno de México a través del Instituto Politécnico Nacional, proyecto SIP 20131702.

Referencias

1. Dagan, I., Glickman, O., Magnini, B. The PASCAL Recognising Textual Entailment Challenge. En: Proc. PASCAL Challenges Workshop on RTE, 2005.
2. Peñas, A., Rodrigo, Á., Sama, V., Verdejo, F. Overview of the answer validation exercise 2006. En: Working Notes, CLEF 2006.
3. Peñas, A., Rodrigo, Á., Verdejo, F. Overview of the answer validation exercise 2007. En: Working Notes, CLEF 2007.
4. Rodrigo, Á., Peñas, A., Verdejo, F. Overview of the answer validation exercise 2008. En: Working Notes, CLEF 2008.
5. Pakray, P., Bandyopadhyay, S., Gelbukh, A. Lexical based two-way RTE System at RTE-5. En: System Report, TAC RTE Notebook, 2009.
6. Pakray, P., Pal, S., Poria, S., Bandyopadhyay, S., Gelbukh, A. JU_CSE_TAC: Textual Entailment Recognition System at TAC RTE-6. En: System Report, Text Analysis Conference Recognizing Textual Entailment Track (TAC RTE) Notebook, 2010.
7. Pakray, P., Barman, U., Bandyopadhyay, U., Gelbukh, A. A Statistics-Based Semantic Textual Entailment System. En: Lecture Notes in Artificial Intelligence, 7094, Springer, 2011, p. 267–276.
8. Pakray, P., Barman, U., Bandyopadhyay, S., Gelbukh, A. Semantic Answer Validation using Universal Networking Language. En: International Journal of Computer Science and Information Technologies, 3(4):4927–4932, 2012.
9. Pakray, P., Bhaskar, P., Banerjee, S., Chandra Pal, B., Bandyopadhyay, S., Gelbukh, A., A Hybrid Question Answering System based on Information Retrieval and Answer Validation. En: CLEF 2011 Workshop on Question Answering For Machine Reading Evaluation (QA4MRE). CLEF 2011 Labs and Workshop. Notebook Papers. Amsterdam. 2011, 16 pp.
10. Bhaskar, P., Pakray, P., Banerjee, S., Banerjee, S., Bandyopadhyay, S., Gelbukh, A. Question Answering System for QA4MRE@CLEF 2012. En: CLEF 2012 Workshop on Question Answering For Machine Reading Evaluation (QA4MRE). CLEF 2012 Labs and Workshop. Notebook Papers, Roma, Italia, 2012.
11. Pakray, P., Gelbukh, A., Bandyopadhyay, S., A Syntactic Textual Entailment System Using Dependency Parser. En: Lecture Notes in Computer Science, N 6008, 2010, Springer, pp. 269–278.
12. Neogi, S., Pakray, P., Bandyopadhyay, S., Gelbukh, A. JU_CSE_NLP: Language Independent Cross-lingual Textual Entailment System. En: *SEM: First Joint Conference on Lexical and Computational Semantics. Collocated with NAACL-HLT, 2012, June 7–8, Montreal, Canadá, 2012.
13. Sidorov, G. Desarrollo de una aplicación para el diálogo en lenguaje natural con un robot móvil. En: Research in computing science, 56, 2012.
14. Sidorov, G. N-gramas sintácticos y su uso en la lingüística computacional, Vectores de investigación, 2013.

Partha Pakray, Alexander Gelbukh y Sivaji Bandyopadhyay

15. Aguilar-Galicia, H.; Sidorov, G.; Ledeneva, Y. Extracción automática de hechos de libros de texto basada en estructuras sintácticas. En: Research in computing science, 55:15–26, 2012.
16. Zhila A., Gelbukh A. Comparison of open information extraction for English and Spanish. En: Computational Linguistics and Intelligent Technologies, 12(19):714–722, 2013.